Curiosity with MARC Data

Mark Schulist

2022-12-21

Study Overview

I will be using my MARC point count data for this part of the project. We collected data by visiting points and counting the birds that we saw there. We visited each point 3 times and would like to know the probability of occupancy of the species in the study. There are 80 points that we visited, and we have data on whether each species was present or not on each visit. Each point can be viewed as a series of three coin flips in a binomial model. I will use Golden-crowned Kinglet as the species of interest in this study.

Data

Here I will import the data. I'm not going to show the entire data set since that would take a lot of space, but it consists of 80x3 matrix where each column corresponds to a point count and each row corresponds to a point.

```
data <- read.table(here("stats/sem_1_final/data.txt")) %>% as.matrix()

colnames(data) <- 1:3
```

Now we are going to use the plausible values method to determine how likely it is to obtain our data given a p value (the long run probability of detecting a Golden-crowned Kinglet across all sites).

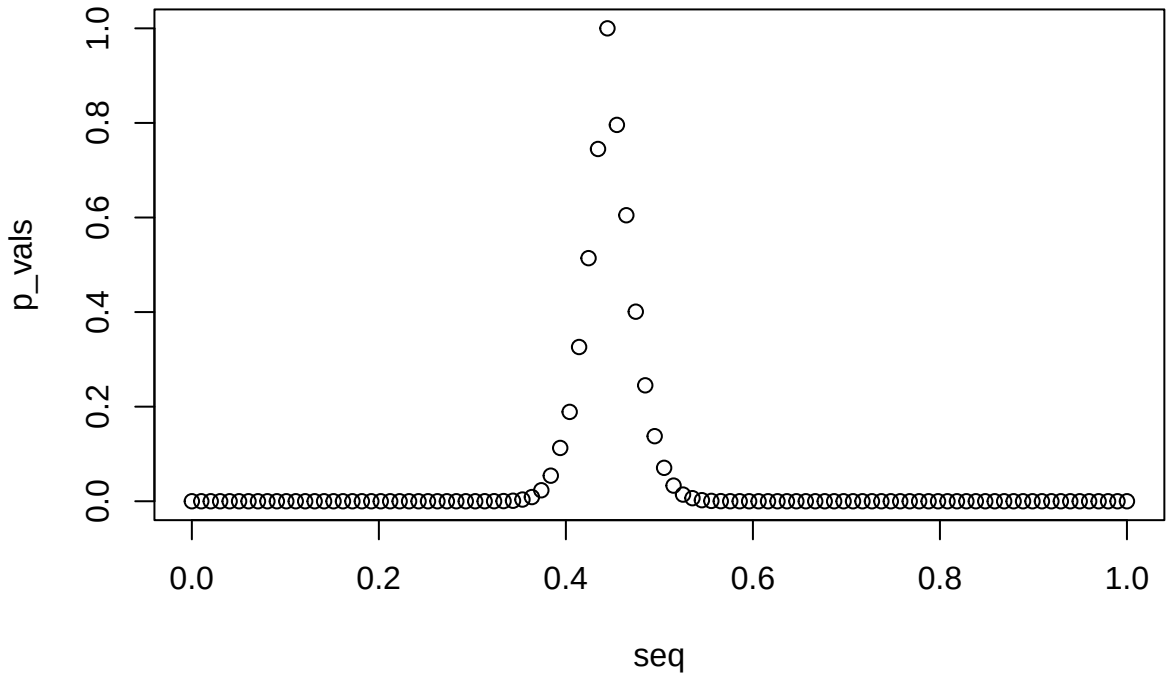
We will use a binomial model to determine the number of heads we get when flipping 240 coins with many possible values of p.

```
seq <- seq(0, 1, length.out = 100)

map(seq, function(x) {
  binom.test(107, 240, p = x, alternative = "two.sided")$p.value
  }
) %>% unlist -> p_vals

names(p_vals) <- seq

plot(seq, p_vals)
```



So now we have the probabilities of getting our number of detections (or more extreme) for many different possible p-values.

If we use 0.05 as a cutoff for p-values, we have a 95% confidence interval of our possible p parameter values.

```
min <- which(p_vals >= 0.05) %>% min()
max <- which(p_vals >= 0.05) %>% max()
paste0("(", round(seq[min], 3), ",", round(seq[max], 3), ")")
```

```
## [1] "(0.384,0.505)"
```

Based on this, I am 95% confident that the true probability of detection of Golden-crowned Kinglet is between 0.384 and 0.505.

Conclusions

The graph above shows the probabilities of seeing our data ($\hat{p} = \frac{107}{240}$) given the value of p. Higher p-values mean that it is more likely that we will see that data, but it does not mean that it is impossible to see the lower p-value parameter values.